A6 **MATH 6205** Due: March 21ˢᵀ, 2022

Upload your solution to Brightspace at the beginning of class on Monday March 21ˢᵗ, 2022.

We consider the synthetic problem of *outlier detection with autoencoders* for the MNIST handwritten digit dataset.

1. Treat digit "0" as the regular class and digit "6" as the outlier class. To generate the respective dataset, import all images from class 0 and add 500 images of class 6 (which is roughly 10% of the total data). Split your synthetic outlier dataset into 90% for training and 10% for testing.

2. Then train two autoencoders using *only the regular images* (i.e. images of the regular class 0) of the training dataset. The first autoencoder is a standard densely connected autoencoder. The second autoencoder is a convolutional autoencoder. Experiment with the architecture for both autoencoders, i.e. vary the number of hidden layers, units per layer, the coding size, etc.

3. Assess the reconstruction loss of both autoencoders on *both the regular and outlier* testing data. Do you observe a higher reconstruction loss for the outlier data for both autoencoders? Are your autoencoder capable of distinguishing regular from outlier images based on the reconstruction loss? If not, give an interpretation of your results.

4. Besides the reconstruction loss, also use the *structural similarity index measure* (SSIM) between the original and reconstructed images as a verification metric. Do your autoencoders achieve worse SSIM (as they should!) on the outlier class? If not, give an interpretation of your results.